



Distance Perception in a loudspeaker-based room auralization system

Favrot, Sylvain Emmanuel; Buchholz, Jörg

Published in:
Audio Engineering Society Convention Papers

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Favrot, S. E., & Buchholz, J. (2009). Distance Perception in a loudspeaker-based room auralization system. In *Audio Engineering Society Convention Papers* (pp. 7854). Praesens Verlag. <http://www.aes.org/e-lib/browse.cfm?elib=15049>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Audio Engineering Society Convention Paper 7854

Presented at the 127th Convention
2009 October 9–12 New York NY, USA

The papers at this Convention have been selected on the basis of a submitted abstract and extended precis that have been peer reviewed by at least two qualified anonymous reviewers. This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Distance perception in loudspeaker-based room auralization

Sylvain Favrot¹, Jörg M. Buchholz¹

¹Centre for Applied Hearing Research, Technical University of Denmark, Kgs. Lyngby, Denmark

Correspondence should be addressed to Sylvain Favrot (sf@elektro.dtu.dk)

ABSTRACT

A loudspeaker-based room auralization (LoRA) system has been recently proposed which efficiently combines modern room acoustic modeling techniques with high-order Ambisonics (HOA) auralization to generate virtual auditory environments (VAEs). The reproduction of the distance of sound events in such VAE is very important for its fidelity. A direct-scaling distance perception experiment was conducted to evaluate the LoRA system including the use of near-field control (NFC) for HOA. Experimental results showed that (i) loudspeaker-based auralization in the LoRA system provides similar distance perception to that of the corresponding real environment and that (ii) NFC-HOA provides a significant increase in the range of perceived distances for near sound sources as compared to standard HOA.

1. INTRODUCTION

Virtual auditory environments (VAEs) have received increased attention in the past decades and are used in a variety of domains from entertainment to psychophysical research. VAEs ideally create the percept of acoustic events that are generated by sound sources in a given space, whereas neither the sound sources nor the space are physically present in the actual listening environment [1]. In order to display sound sources at specific positions in the virtual space, understanding of human auditory localization is required. Whereas the perception of the direction

of the source has been widely studied [1], much less is known about auditory distance perception. Besides the fact that reproduction of the distance of sound sources in a VAE is very important for its fidelity, VAEs provide a tool to systematically study the influence of individual distance cues (e.g. [2]) and thus to better understand the mechanisms underlying distance perception.

A change of the physical distance between a listener and a sound source induces the change of four main characteristics of the acoustical waveforms at the lis-

tener's ears: intensity, direct-to-reverberant energy ratio, binaural differences (for sources in the acoustic near-field) and spectrum [2]. The intensity and the spectrum are ambiguous cues since their variation can also result from a change of the sound source characteristics (e.g., the input signal to a loudspeaker source), whereas the direct-to-reverberant ratio and the binaural differences are independent of the source characteristics. This implies that a realistic reproduction of the direct-to-reverberant ratio and the binaural differences (for near-field sources) potentially allows for an independent control of the distance and intensity of a source in a VAE. This is of particular importance in applications such as auditory displays where the distance can code the importance of an event.

Strategies for accurately reproducing these cues in a VAE are closely linked to the playback technique: either binaural or loudspeaker-based. This study focuses on the ability of loudspeaker-based systems to provide realistic distance perception for both near- and far-field sources in various acoustic environments. The loudspeaker-based room auralization (LoRA) system [3] is here taken as an example to demonstrate that realistic distance perception can be achieved in loudspeaker-based VAEs. The LoRA system efficiently combines modern room acoustic modeling techniques with high-order Ambisonics (HOA) auralization.

The reproduction of the different distance cues to listeners in VAEs is of various degree of complexity. While the intensity cue and the spectrum cues are merely controlled by the overall playback gain in different frequency channels, the reproduction of natural (i) direct-to-reverberant ratio cue and (ii) binaural differences cue for near-field sources with loudspeakers requires more technical effort.

Accuracy of the provided direct-to-reverberant ratio cues in the VAE are determined by the precision of the different components in the overall chain of the LoRA system, i.e. the acoustic room model providing the room impulse response (RIR), the HOA auralization of the early part of the (RIR) and the reverberation algorithm for the late part of the RIR. The goal of the first part of this study is to assess the quality or realism of this cue provided by VAEs such as the LoRA system. For this purpose, subjective distance estimations in VAEs need to be com-

pared to estimations in corresponding real environments. Since distance perception is strongly influenced by visual cues, an appropriate auditory-only reference condition needs to be defined. Individualized binaural room impulse responses (BRIRs) have been previously used for studying distance perception in real environments [2] and seem to provide a suitable auditory-only reference condition. However, this technique requires a substantial amount of time and it has been shown that auditory distance perception is not significantly modified by the use of non-individualized BRIRs in reverberant rooms for distances between 0.3 and 13.8 m [4]. Thus, dummy-head recorded BRIRs provided the reference condition for the present distance perception experiments.

Binaural differences are an important cue for distance perception of nearby sources which, unlike binaural techniques, are challenging to render with a loudspeaker array as signals at the listener ears are not precisely controlled. Ambisonics is a commonly used spatialization technique which primarily reproduces far-field sources (i.e. producing plane waves). The near-field control (NFC) method [5] has been proposed to reproduce spherical wave sound fields produced by near-field sources, typically closer than the loudspeaker array radius, and has been evaluated physically. The second part of this study provides a perceptual evaluation of the distance of near-field sources rendered with the NFC-HOA technique in both anechoic and reverberant environments in the framework of the LoRA system. Whereas the reverberant condition reflects the most common case in VAEs, the anechoic condition is expected to provide the most sensitive case for studying the influence of the NFC technique.

2. METHODS

In this study, the perception of distance was assessed by using a direct+scaling method in environments restricted to the auditory modality. Both the fidelity of distance perception in loudspeaker-generated VAEs and the effect of the NFC-HOA method on the reproduction of near-field sources were investigated.

The experiments took place in an acoustically damped room where a spherical 3D array of 29 loudspeakers with a radius of 1.8 m was used. Seven

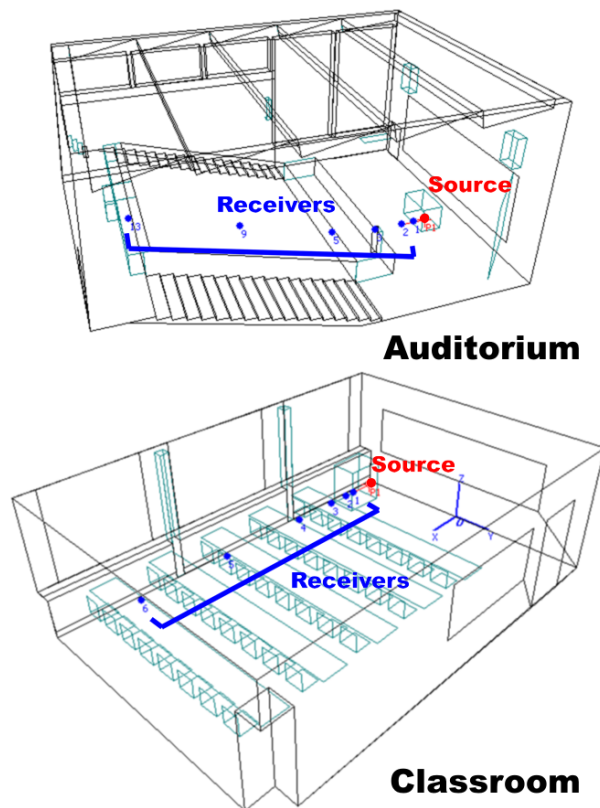


Fig. 1: 3D models of the rooms and positions of the sources and receivers used in the experiment.

normal-hearing persons participated in the experiment.

2.1. Stimuli

2.1.1. Acoustic environments

Three rooms were considered in this experiment: a classroom, a large auditorium and an anechoic room (Fig. 1). In each room, a set of receivers positioned at various distances (see Tab. 1) from an omnidirectional source was selected. The tested conditions are listed in Tab. 2. Conditions C1 to C8 refer to the investigation of distance perception in LoRA generated VAEs and conditions C9 to C15 to the NFC investigation for which the range of distances was restricted from 0.6 to 4.3 m.

2.1.2. Binaural recordings for the reference conditions

For conditions C5 to C8 and C15, binaural room im-

Rooms	Distances [m]					
Classroom	0.6	0.8	1.3	2.3	4.3	6.3
Auditorium	0.5	1.0	2.2	4.0	7.5	11.2
Anechoic	0.6	0.8	1.3	2.3	4.3	6.3

Table 1: Table of the source-receiver distances used in the experiment.

pulse responses (BRIRs) were recorded with a head and torso simulator (HATS) from Brüel & Kjær at the receiver positions defined in Fig. 1 in the actual physical rooms. BRIRs have also been recorded in an anechoic room (C15) for comparison to the anechoic NFC-HOA conditions. The non-individualized BRIRs measured in the anechoic room are known to greatly degrade distance perception and typically result in in-the-head-localization [1]. However, this anechoic condition was included to allow for a comparison between non-individual tailored binaural and loudspeaker-based systems. The source was placed in the 90° azimuth direction of the receiver to provide significant binaural differences. The same condition at 0° azimuth was not tested because a preliminary test showed that the stimuli at different distances were not discriminable.

2.1.3. Room simulation

For condition C1 to C4, the classroom and the auditorium were simulated with ODEON (Fig. 1), a room acoustic modeling software, where the source and receiver positions were set according to Fig. 1. For each source-receiver configuration, room impulse responses (RIRs) were derived from the ODEON software. These RIRs were processed by the LoRA system [3]. They were first decomposed into an early discrete part (i.e., the direct sound and the early reflections) and a late diffuse part. For these two rooms, the discrete part of the room's response was auralized with fourth-order 3D HOA and the diffuse reverberation part was auralized by multiplying intensity envelopes of the room's response with noise that is uncorrelated across loudspeakers. A set of multi-channel RIRs (mRIRs) was thus obtained for the classroom and the large auditorium.

2.1.4. Near-field control with high-order Ambisonics

For conditions C9 to C12, another set of mRIRs was derived by integrating near-field control (NFC) fil-

	Auralization technique	Room	Az.	Process
C1	LoRA	Class.	0°	IN
C2	LoRA	Class.	0°	
C3	LoRA	Aud.	0°	
C4	LoRA	Aud.	0°	
C5	Binaural	Class.	0°	IN
C6	Binaural	Class.	0°	
C7	Binaural	Aud.	0°	IN
C8	Binaural	Aud.	0°	
C9	LoRA	Class.	0°	NFC+IN
C10	LoRA	Class.	90°	NFC+IN
C11	HOA 4 th 3D	Anec.	0°	NFC+IN
C12	HOA 4 th 3D	Anec.	90°	NFC+IN
C13	HOA 7 th 2D	Anec.	0°	NFC+IN
C14	HOA 7 th 2D	Anec.	90°	NFC+IN
C15	Binaural	Anec.	90°	IN

Table 2: Test conditions are listed with the auralization technique, the room (classroom, auditorium or anechoic), the azimuth of the direction of the source (Az.) and with either no processing, intensity normalization (IN) or NFC and intensity normalization (NFC+IN).

ters to HOA [5]. This technique consists of filtering the Ambisonic encoded components according to the distance of the source (ρ) and the radius of the loudspeaker array (R) to represent finite distance sources with HOA. The formula describing these filters, $H_m^{NFC(\rho,R)}(\omega)$, is shown in equation (10) in [5] and their frequency responses are shown in Fig. 2 for each Ambisonic component m equals 0 to 7 for a source of 0.5 m. These filters result in a very large bass-boost when reproducing very close sources. A Thikonov regularization, inspired by [6], was used to limit the amplification at 30dB. Regularized filters were described by:

$$H_{reg} = \frac{H_m^{NFC(\rho,R)} \cdot \lambda^2}{H_m^{NFC(\rho,R)^2} + \lambda^2} \quad (1)$$

with $\lambda = 10^{30/20}$ and are plotted (dashed curved) in Fig. 2.

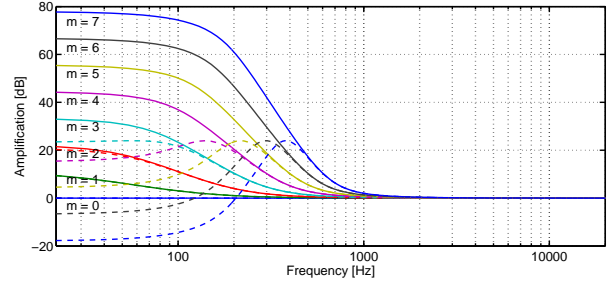


Fig. 2: Frequency response of the near-field coding (NFC) and the regularized NFC (dashed curves) filters.

An anechoic room was simulated by selecting the direct sound alone from the classroom RIRs for each source-receiver configurations. The anechoic room and the classroom RIRs were auralized with both fourth-order 3D and seventh-order 2D Ambisonics with NFC filters for an incoming direction of 0 and 90 degree azimuth.

2.1.5. Calibration

The mRIRs were obtained for all rooms with a source sound power set at 72 dB re. 1W in ODEON (corresponding to 61 dB SPL at 1 m in free field). For each room, the recorded BRIRs were adjusted such that for a source-receiver distance of 2.3 m, the computed SPL in the 500 Hz octave band derived in ODEON matches the SPL in the 500 Hz octave band of the BRIR.

Intensity normalization was performed for the conditions specified in Tab. 2 and relied on the SPL level in the 500 Hz octave band where the maximum energy of the speech signals used in this experiment was found. For each room, the normalization gains for the mRIRs were obtained by subtracting the 500 Hz octave band SPL of the RIRs derived by ODEON for the 2.3 m configuration to the one for the considered distance configuration. Normalization gains for the BRIRs were similarly obtained by deriving the 500 Hz octave band SPLs for the 0° azimuth configuration.

2.1.6. Playback

Obtained binaural and multichannel responses were convolved with the same anechoic Danish speech sentence (approximately 2 s long) for all distances of the same condition. Convolved sentences for mul-

tichannel conditions were played on the 29 3D loudspeaker array or the restricted 16-loudspeaker horizontal array for the HOA seventh 2D conditions (C13 and C14). A loudspeaker equalization was performed to flatten each of the loudspeaker frequency responses recorded at the center of the array. The binaural convolved sentences were played back on headphones that were equalized to flatten the average frequency response between the headphone and the ears of the HATS for different headphone positions.

2.2. Procedure

Each condition was divided into three blocks consisting of three repetitions of each distance configuration presented in a random order. To avoid any interference by visual cues provided by the playback room, the listeners were unfamiliar to the laboratory area and were blindfolded before entering the playback room. Subjects were seated on a chair at the center of the loudspeaker array and asked to remain still during each block. In the beginning of each block, the auralized sentences corresponding to the closest and furthest distance for this condition were played. It provided the subject with the distance range of the sound events together with the distance variation cue in the present block. The value of the actual physical distance range was not revealed to the subject. Then, after each auralized sentence, subjects were asked to answer aurally the question: “How far away does the sound appear?” with a distance in meter. Each subject performed a short training session beforehand in order to be familiarized with the task.

3. RESULTS

3.1. Overall data and power-law fitting

Means and standard deviations of the logarithmic transform of the apparent distance across repetitions are plotted in Fig. 3 for each subject and all conditions. In each panel, a small variation was introduced in the physical distance in order to avoid occlusion of symbols for like responses across subjects.

Standard deviation across repetitions were ranging from 0 to 0.35 for all subjects, conditions and distances except for subject ‘S’ who showed a range between 0.12 and 0.66 for the binaural conditions. For all conditions, these within-subjects variations

were small in comparison to the large inter-subject variability of the data shown in Fig. 3. In order to model the data, the apparent distances were fitted with a simple linear function in the double logarithmic domain, which in the linear domain corresponds to a power-law fitting $r' = kr^a$, where r and r' represent the physical and apparent distances in meters respectively, a the power-law exponent and k a constant. The fitted curves for an example subject (‘T’) were indicated in Fig. 3 by a blue solid line. For conditions when all natural cues were presented (i.e., conditions without intensity normalization: C1, C3, C5 and C7), the average value of a across subjects was 0.52 (s.d. 0.16) and always lower than its veridical value of 1. This highlights the compressive behavior of the distance perception as seen in various auditory studies reviewed in [7]. The constant k represents the offset in the absolute estimation of distances and was ranging, for these conditions, from 0.5 to 4.7 m for the classroom and from 0.6 to 11.1 m in the auditorium, reflecting the above mentioned large inter-subject variability.

3.2. Distance perception with the LoRA system

Power-law exponents, i.e. the degree of compression of the fitted power function, for all subjects for the first part of the experiment (i.e., conditions C1 to C8) showed significant positive values. The mean values and confidence intervals of these exponents across subjects are shown in Fig. 4 and the lowest value of the confidence intervals was 0.16. All subjects were thus able to perceive a change of distance even when the intensity of the stimuli was normalized with distance (IN), indicating the usefulness of the direct-to-reverberation cue provided by both auralization techniques. A further analyze showed that the exponent a of the natural intensity conditions was significantly larger than the one of the corresponding intensity normalized conditions for the classroom and the auditorium for both auralization techniques (approximatively by a value of 0.2). As expected, when both intensity and the direct-to-reverberant ratio cue are provided, subjects had a greater distance discrimination compared to when only the later cue was provided.

When comparing distance perception between both auralization techniques, mean values of the difference of exponent a and constant $\log(k)$ between the

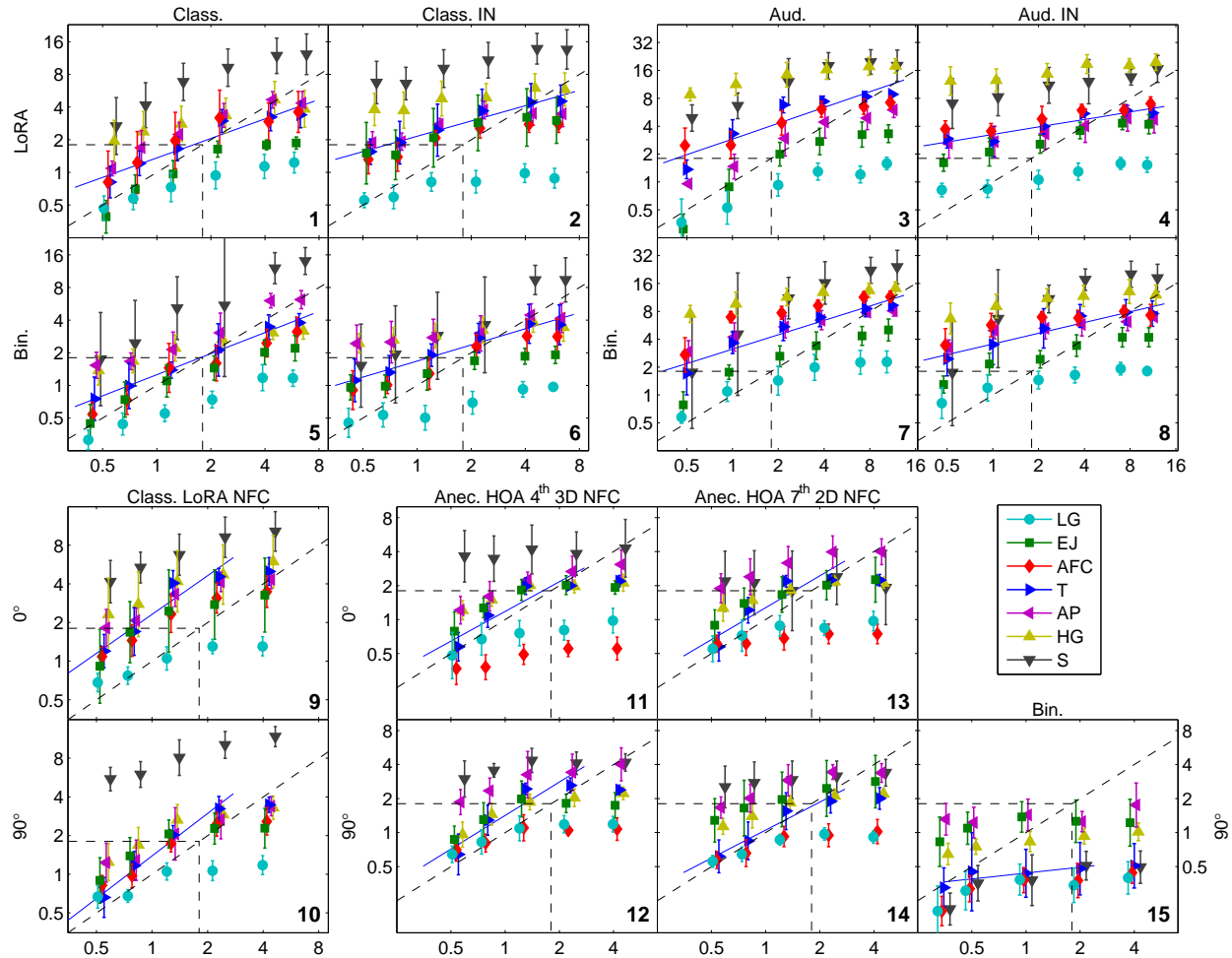


Fig. 3: Mean values and standard deviations of the logarithmic apparent distance plotted against the logarithmic physical distance for all conditions. Axis labels units are in meters. The dashed vertical and horizontal lines represent the loudspeaker array radius and the diagonal dashed line, the physical-estimated identity. The condition number is written on the bottom right corner of each plot.

LoRA and the binaural conditions were not significantly different from 0 for both rooms, with or without intensity normalization. Hence, the experiment show no significant difference in quality or realism in distance perception between the room simulation based LoRA system and the dummy-head recording-based binaural system.

3.3. Distance perception of near-field sources

In order to evaluate the effect of the NFC method on HOA for near sound sources, i.e. for sources within the loudspeaker array of 1.8 m radius, only

the data corresponding to physical distances smaller than 2.3 m were considered in the power-law data fitting. In the anechoic environment, the means of the power-law exponents across subjects for conditions when the distance was simulated with NFC filters only (C11 to C14) were significantly positive (see Fig. 5) and showed an average value across these four conditions of 0.45 which is comparable to the average exponent a of 0.52 for the natural intensity condition in reverberant environments (i.e., conditions C1, C3, C5 and C7). Sources were thus perceived from within the loudspeaker array when

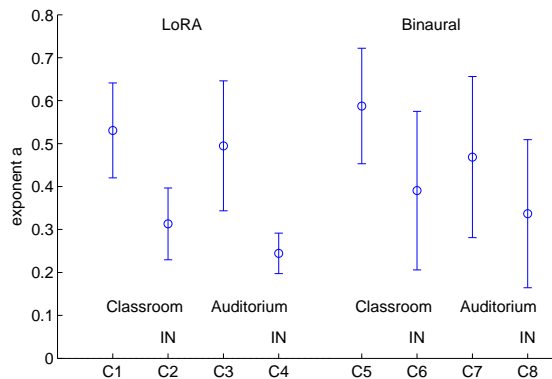


Fig. 4: Mean values and confidence intervals of the fitted power-law exponent a for conditions C1 to C8.

reproduced with the near-field control method for HOA. However, absolute distance estimations varied greatly across subjects (average k of 1.6 m, s. d. 1 m) as shown in Fig. 3 indicating that sources closer than the loudspeaker array radius (1.8 m) were not always estimated below 1.8 m.

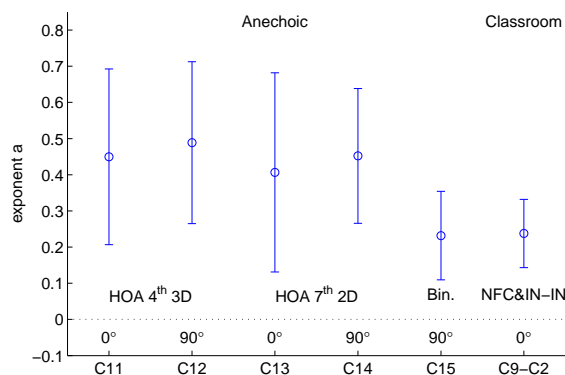


Fig. 5: Mean values and confidence intervals of the fitted power-law exponent a for conditions C11 to C15 and C9-C2.

The mean values of the difference of exponent a between conditions auralized with HOA fourth order 3D and HOA seventh order 2D were not significantly different from zero. The direction of the source did also not have any influence on the exponent.

The mean value of the exponent a for the binau-

ally auralized condition (C15) in the anechoic environment was of 0.23 (see Fig. 5), which was lower than the mean for the NFC-HOA auralized conditions (C12 and C14, average a value of 0.47). However, the mean of the differences (C12-C15 and C14-C15) was not significantly different from zero. The value of the average constant k was 0.6 m, significantly lower than the two loudspeaker-based conditions (C12 and C14, average k value of 1.5 m). Hence, distances were globally underestimated with the binaural technique auralization compared to the loudspeaker-based one. This might be due to the limitation of the use of non-individual BRIRs in anechoic environments that often leads to in-the-head-localization [1].

In the classroom with intensity normalized stimuli, apparent distances showed a larger power-law exponent when the NFC method was used: the mean of the difference of a was of 0.24 (See Fig. 5, C9-C2).

4. DISCUSSION

The inter-subject variability in distance estimations was very large in all experimental conditions, an aspect that has previously been pointed out by other studies (e.g., [8, 7]), highlighting the rather poor auditory sensitivity to distance. The large variability might have been further influenced by the fact that evaluation of distances based solely on auditory cues was not a familiar task for the test-subjects and that the direct-scaling method was used without prior reference in meter to avoid bias of the responses.

The study showed significant different degrees of compression of the fitted power function for different conditions. First, in reverberant rooms, the power-law exponent was decreased when the stimuli intensity was normalized with distance for both binaural and loudspeaker-based auralization techniques (by approximatively 0.2). Second, the power-law exponent was increased when near-field compensation filters were applied to HOA auralization (by approximatively 0.24). This showed that the removing of the intensity cue decreased the range of estimated distances and that NFC significantly contributes to distance perception in HOA-based room auralization respectively.

In anechoic environment, NFC-HOA (at least for fourth- and seventh-order) auralized sources positioned as close as 0.6 m and within the loudspeaker

array radius were perceived closer (average a of 0.45) compared to the same source auralized without NFC. This demonstrates the ability of rendering sources within the loudspeaker array with NFC-HOA.

The experimental results showed no significant differences between the binaural recording-based and the loudspeaker-based (LoRA system) auralization technique both in terms of the power-law exponent a and the absolute values k . This was the case when all natural cues were presented as well as when the intensity of the stimuli was normalized with distance. Loudspeaker-based auralization in the LoRA system seems thus able to reproduce auditory environments in which distance perception is similar to that of the corresponding real environment. However, one could argue that the experiment was not able to show a significant effect between the two auralization techniques due to limitations in the statistical relevance of the study. In this regards, the sensitivity of the experiment could be improved by increasing the number of subjects and repetitions. Nevertheless, even if an effect on distance perception between the two auralization techniques exists, it is expected to be smaller than the effect of removing the intensity cue or applying NFC, which the experiment provided evidence for.

5. CONCLUSION

The present auditory distance perception experiment provided evidence for the following points:

- Loudspeaker-based auralization in the LoRA system provides similar distance perception to that of the corresponding real environment,
- NFC-HOA provides a significant increase in the range of perceived distances for near sound sources as compared to standard HOA,
- NFC-HOA auralization allows for rendering sources within the loudspeaker array both in anechoic and reverberant environments,
- The use of non-individual BRIRs led to underestimated distances compared to the use of NFC-HOA.

6. REFERENCES

- [1] J. Blauert, "Spatial hearing: the psychophysics of human sound localization," MIT Press, third ed., 2001.
- [2] P. Zahorik, "Assessing auditory distance perception using virtual acoustics," *Journal of the Acoustical Society of America*, **111**, 2002.
- [3] S. Favrot and J. Buchholz, "LoRA - A loudspeaker-based room auralisation system," *Acta Acustica*, submitted, 2009.
- [4] P. Zahorik, "Auditory display of sound source distance," presented at the ICAD, Kyoto, Japan, 2002.
- [5] J. Daniel, "Spatial sound encoding including near field effect: introducing distance coding filters and a viable new Ambisonic format," presented at the AES 23rd conference, Copenhagen, Denmark, 2003.
- [6] S. Moreau, J. Daniel and S. Bertet, "3D sound field recording with HOA - Objective measurements and validation of spherical microphone," presented at the AES 120th convention, Paris, France, 2006.
- [7] P. Zahorik, "Auditory distance perception in humans: a summary of past and present research," *Acta Acustica*, **91**, 2005.
- [8] S. Nielsen, "Auditory distance perception in different rooms," *Journal of the Audio Engineering Society*, **41**, 1993.